The Management of Data: Preliminary Research Results

Dale Goodhue
Judith Quillard
John F. Rockart

90s: 86-022

Working Paper

Management in the 1990s

Massachusetts Institute of Technology
Sloan School of Management

The Management of Data:
Preliminary Research Results

Dale Goodhue
Judith Quillard
John F. Rockart

90s: 86-022

May, 1986

CISR WP#140
Sloan WP #1786-86

Management in the 1990s
Sloan School of Management
Massachusetts Institute of Technology

Management in the 1990s is an industry and governmental agency supported research program. Its aim is to develop a better understanding of the managerial issues of the 1990s and how to deal most effectively with them, particularly as these issues revolve around anticipated advances in Information Technology.

Assisting the work of the Sloan School scholars with financial support and as working partners in research are:

American Express Travel Related Services Company
Arthur Young and Company
British Petroleum Company, p.l.c.
BellSouth Corporation
Digital Equipment Corporation
Eastman Kodak Company
General Motors Corporation
International Computers, Ltd.
MCI Communications Corporation
United States Internal Revenue Service

THE MANAGEMENT OF DATA
PRELIMINARY RESEARCH RESULTS

1.0 INTRODUCTION

Today, corporations are placing increasing emphasis on the management of data. A number of factors, which include a significantly more competitive environment, more ubiquitous personal computers, rapidly improving software, and more computer-literate personnel, have led to a demand for more and better data. Acknowledging this, more and more companies are trying to follow the exhortations of authors such as Diebold [1979], Edelman [1981], and Appleton [1986] to "manage the data resource". These, and other authors argue that information and data play a significant and growing role in achieving organizational performance and, therefore, must be managed proactively.

The available literature, however, contains little to suggest that organizations have successful track records in managing data from a business perspective, rather than a technical viewpoint. In order to learn more about the characteristics of effective approaches to "managing the data resource", we studied 17 data management efforts in a set of diverse companies. Almost all of these efforts were viewed as successful by both information systems staff and user management. Our research focused on the managerial motivations, planning processes, and types of outputs achieved.

The authors would like to thank Judith Green Carmody, William A. Gillette, Lisa A. Johnson, Ruth E. Kocher, John S. Lively, and Judy E. Strauss, who did their Master's thesis research in conjunction with this study.

Three major themes emerged from our study. They are:

- o The management of data is becoming increasingly important. Major business and technology trends suggest that, if anything, this importance will increase.
- o There is no single, dominant approach to improving the management of data. Rather, firms have adopted multiple approaches, contingent upon business needs.
- o A set of fundamental managerial issues must be considered in selecting the appropriate approach to data management.

The rest of Section 1 describes the research method, the growing importance of data management, and some current approaches to managing data in the literature. Section 2 presents a framework and describes the choices we saw. Section 3 summarizes the framework. Section 4 then addresses several key managerial issues that managers should explicitly address as they consider the options suggested by our framework. Section 5 concludes the paper by emphasizing the several major themes in our results.

1.1 Research Approach

These themes resulted from the first phase of a data management study being conducted by the Center for Information Systems Research (CISR) at MIT's Sloan School of Management. Case studies, each involving 4 to 6 days of interviews, were conducted in six large corportions during the spring of 1985. In addition, CISR reseachers spent 1/2 to 3 days in each of 5 other organizations to learn about their data management efforts. The companies, summarized in Figure 1 with disguised names, represent a range of industries, including electronics, consumer goods, insurance, and energy. In each firm, we focused on one to four data management programs or projects.

These firms discussed with us the data-related policies, processes, controls, standards, and tools that were in place (or had been tried). Also discussed were the factors which motivated each organization to take action, and the results which were achieved. In addition, our interviews collected the opinions of I/S managers, line managers, and staff professionals, regarding the most important problems and issues concerning the management and use of data.

Figure 1. Firms in the CISR Data Management Study

Disguised Name	Description	Data Management efforts cited in this paper
Blaine Corporation	Personal care products; Fortune 500	Data access services for end users
Crockett	Canadian subsidiary of Fortune 500 U.S. computer company	Integrated database for finance and administrative applications and managerial reporting
Diverse Conglomerate	Fortune 500	Information database for use by senior management
Dobbs Insurance	Among Top 10 in assets	Strategic data planning and new systems for one division
Foothill Computer	Fortune 250	1. Information centers provide data consulting 2. Set of standard data definitions established by corporate task force 3. Strategic data planning used for order flow function
Global Products, Inc.	Fortune 500 manufacturing	Subject area databases being gradually implemented
Matrac Corporation	Fortune 100 manufacturing	Data access services for end users
Sierra Energy	Among top 10 energy companies	Information database for financial reporting
Spectrum Electronics	Fortune 100	Integrated manufacturing database in one division
Waverly Chemicals	Fortune 500 diversified chemical company	1. Common manufacturing systems largest division 2. Common accounting systems used by multiple divisions 3. Strategic data modeling done for I/S planning in largest division 4. Set of standard data definitions established by corporate group
Windsor Products	Fortune 100	1. Several information databases built 2. "Data Charting" effort

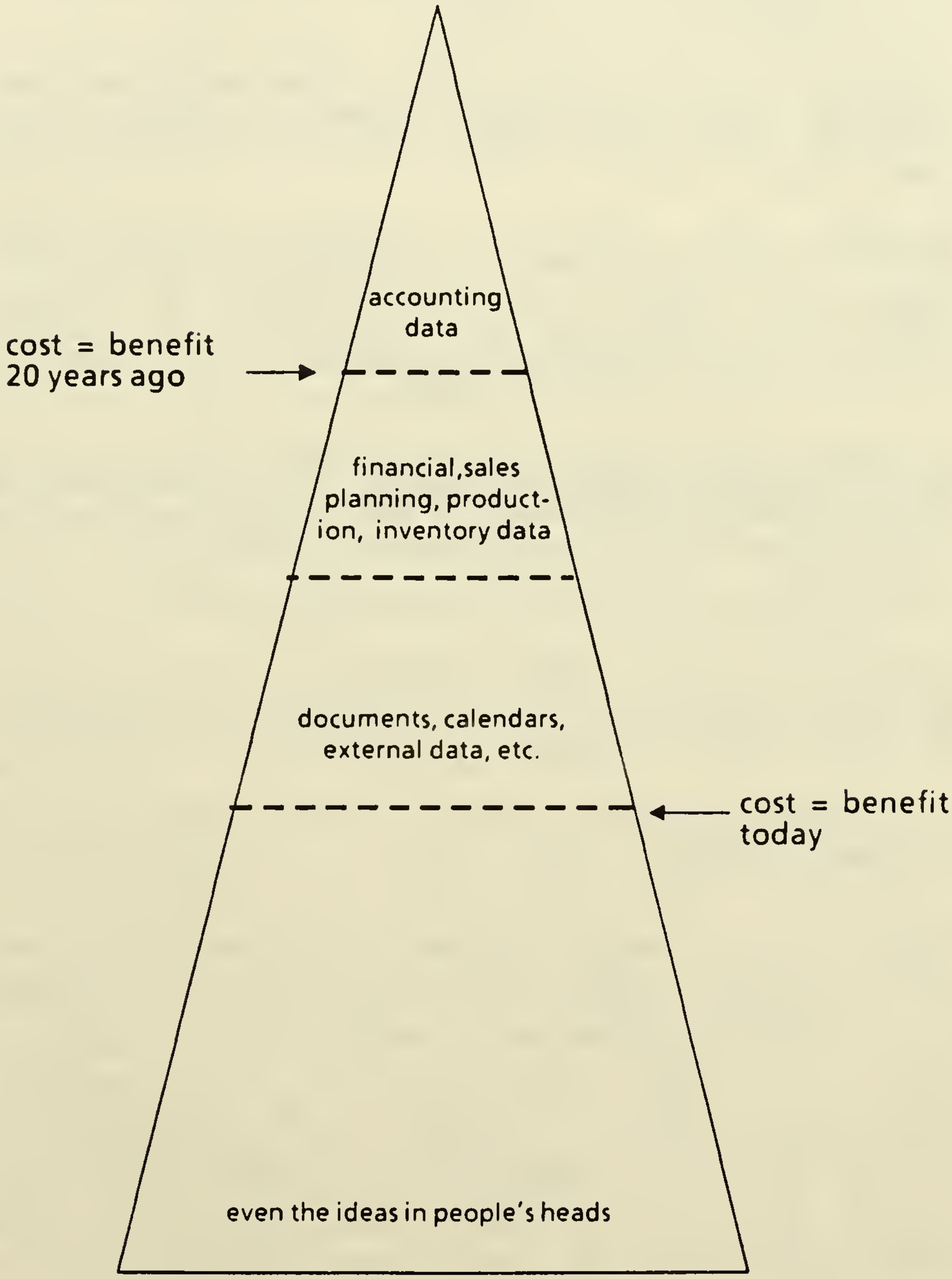
1.2 Managing Data Is Becoming Increasingly Important

A rationale for the growing importance of data management is found in Huber's [1984] prediction that "Post industrial society will be characterized by more and increasing knowledge, more and increasing complexity, and more and increasing turbulance" (author's emphasis). These environmental changes, Huber suggests, will require the management of all organizations to have up-to-date information from multiple sources at many levels of aggregation. Thus, organizations, to survive, will need improved designs for information acquisition and distribution.

As the business need has increased, so has the technical capability to deliver information. Over the last twenty years, there have been dramatic declines in hardware costs and gains in software functionality. Thus, as the benefits to be gained from timely, accurate information have increased, the costs have decreased. As illustrated in Figure 2, it has become cost effective for organizations to manage new types and greater quantities of information with computer technology.

Because a rapidly increasing amount of data is now available in electronic form, managing that data effectively is growing not only in importance, but also in complexity. Realistically, corporate attempts to manage the data resource do not start from a clean slate. An underlying problem with managing data in organizations today is the difficulty of integrating data originally designed to meet the needs of isolated applications, which were often developed in widely dispersed, almost autonomous, sub-organizations within a large corporation. This problem of decentralized development has led to discrepancies in data definitions, accuracy differences, and timing and coordination problems. The resulting lack of accessible, quality data today often limits managerial and staff access to information, hinders system development and maintenance activities, and even constrains the ability to undertake strategic initiatives.

Figure 2. HOW MUCH INFORMATION IS IT COST-BENEFICIAL TO "MANAGE"



Due to decreasing costs and increasing benefits, the sets of data that have been computerized in most corporations have expanded from accounting data to include a complete range of operational activities, staff functions, and line managerial activities. In addition to numerical data, increasingly, text is being stored. Efforts are underway to store, through "expert systems", some of the knowledge previously available only in people's heads. That much more data will be stored and accessed electronically in the next decade is relatively certain.

1.3 Current Approaches to Data Management

What means should firms use to achieve the goal of managing data in a manner which best contributes to business objectives? Various approaches which have been discussed in the existing data resource management (DRM) literature can be categorized as follows:

- o Approaches with a technical focus. These include tools and techniques such as database management systems, data dictionaries [Ross, 1981], data entity relationship modeling [Chen, 1976].

- o Approaches with a focus on organizational responsibilities. These include the establishment of organizational units such as database administration and data administration [Tillman, 1984], [Voell, 1979], and the formulation of administrative policies and procedures covering areas such as data ownership, access, and security [Appleton, 1984].

- o Approaches with a focus on business-related planning. These include planning processes and methods such as Strategic Data Planning [Martin, 1982] and BSP [IBM, 1981] that link the acquisition and use of data with business objectives.

It is increasingly evident that no one category provides a completely adequate approach. Coulson [1982] acknowledges that many efforts to straighten out data management problems through the implementation of a data dictionary have failed. Kahn [1983] presents empirical evidence suggesting most data administration groups have had little or no success in correcting key data management problems.

Because the ultimate goal is not to put in place tools or to create organizational units, but rather to link data needs with the needs of the business, much attention has focused on the third type of approach. These planning approaches, however, require significant resource commitments, and are often not easy to undertake. For example, Zmud [1983] notes that "to be done well, BSP requires many organizational members to make considerable investments of time and effort" (Chapter 10, p.281). Section 2.2.1 will discuss some reasons why large-scale, top-down strategic data planning efforts are difficult to accomplish.

2.0 A CONTINGENCY APPROACH TO DATA MANAGEMENT

The picture which emerges from our case studies is that effective management efforts fit no single clear pattern. Rather there are several "paths" to improving the management of data. Which path is selected depends heavily on organizational considerations. Successful efforts are very diverse in business motivation, organizational scope, planning method, and type of result obtained. Certainly one must be careful in generalizing from less than twenty examples, but this variety (or contingency approach) in our case studies seems also to be borne out in discussions which we have had with a number of other companies.

In analyzing the cases, we see four key areas which represent an interlinked set of choices that organizations make as they undertake a data management effort. These areas are:

- The identification of a <u>business objective</u>. In the successful companies in our sample, data management actions were almost always justified not by conceptual or technical arguments, but by one of four compelling business needs: coordination, flexibility, improved managerial information, or I/S efficiency.

- The <u>scope</u> of the data management project. The firms we studied had explicitly defined and limited the scope of their efforts. Some focused on a functional area (such as finance), others on a division, while some were corporate-wide.

- The <u>data planning</u> method. Top-down, in-depth strategic data modeling was not the only data planning process we saw. In fact, there appear to be a number of obstacles in accomplishing a large-scale strategic data planning effort. The planning processes utilized in our cases varied widely in terms of their formality, their detail, and their emphasis on data models. The range of options varied from strategic data modeling to more limited planning approaches to no planning whatsoever.

- The <u>"product"</u> of the data management effort. Much of the existing DRM literature centers on the implementation of subject area databases [Martin, 1982]. In our case studies, however, we saw five distinct "products", which were the end results of the data management project team's work. These products are: subject area databases for operational systems, common systems, information databases, data access services, and architectural foundations for future systems.

Figure 3 presents these four components and shows the choices generalized from our cases. This simple framework represents a starting point for visualizing what organizations are actually doing with respect to data management. The next four subsections of the paper (2.1 - 2.4) discuss each of the elements of the framework. We begin with the most tangible, the "products", and then in turn discuss planning processes, scope, and business objectives.

2.1 Five Data Management "Products"

The "product" most common in the existing DRM literature is a set of subject area databases used by multiple operational systems. We also identified as products common systems, and what we will call information databases. These three products are similar in that each involves new systems development. In addition to these "new system" products, we identified two other products: data access services, and architectural foundations for the future. This section discusses at some length each of the products and presents case examples.

2.1.1 Subject Area Databases for Operational Systems

Subject area databases (SADBs) contain data which is organized around important business entities or subject areas, such as customers and products, rather than around individual applications, such as order processing or manufacturing scheduling. Many different operational applications may share (both access and update) data from a single set of SADBs. In four of the companies we studied, the results can best be described as SADBs for operational data. In the first, strategic data planning was the basis for the effort.

Dobbs Insurance Companies is one of the nation's largest insurance companies. Dobbs has five major divisions which operate relatively autonomously. In the late 1970s, the Group Insurance Division (GID) began to address issues of data

Figure 3. Framework of Data Management Choices

Business Objective	Scope	Data Planning Process	Product
• Coordination	• Corporate	• Strategic Data Modeling	• Common Systems
• Flexibility	• Division	• "80/20" Methods	• SADBs (operational data)
• Improved Managerial Information	• Function	• Targeting	• Information Databases
• I/S Efficiency		• None	• Data Access Services
			• Architectural Foundations

resource management. Due to a changing business environment and a recent reorganization, the division wanted to reevaluate its operational systems. After a significant amount of preliminary work, GID undertook a strategic data planning effort based on James Martin's approach. From the resulting strategic data model, a dozen subject area databases (including customer, asset, and product) have been implemented as the foundation for the division's applications systems.

Subject area databases for operational systems can also be designed and implemented more gradually.

Global Products, Inc., had a history of using selected reference files for static data elements such as customers, products, prices, vendors, etc. Integration of application systems was achieved by passing transaction data elements from program to program following a predefined flow. In the mid-1970s, Global Products decided to replace this sequential, application-oriented approach with a subject area database approach for its finance, distribution and marketing functions. The goal was achieved not by forcing rewrites of all existing systems, but by offering new or rewritten applications access to the SADBs. The approach resulted in a growing set of integrated subject area databases. This improved the accuracy and timeliness of operational data, and reduced system development and maintenance effort.

The difficulty with such an incremental approach is that each new system's data requirements may not be consistent with the current design of the SADBs. Where necessary, the set of databases must be revised to incorporate the requirements of the new system. The advantage, however, is that the changeover can be accomplished with minimum disruption to existing systems and a preplanned movement toward a consistent set of databases can occur, taking into account the resource constraints of the organization. Today, approximately 40-50 databases are shared by 30 applications at Global Products.

Data management efforts began in the Consumer Products Division of Spectrum Electronics more than ten years ago with the creation of a centralized manufacturing database for its twelve plants. Pressed by Asian competition, the Consumer Products Division, with annual sales of almost $1 billion, started by developing and implementing a logical data model for five key manufacturing applications. The division then proceeded to incorporate all remaining manufacturing systems into the database, and now has subject area data for vendors, parts, assemblies, etc.

2.1.2 Common Systems

A second type of data management "product" is the operational data files or databases which are developed for common systems. Common systems are applications developed by a single, most often a central, organization to be used by multiple organizational units. Physically, there can be one or multiple copies of the system. As opposed to the subject area databases discussed above, which are developed to be shared by a range of applications, a common systems approach tends to focus on replacing existing systems in one specific application area.

Many firms already have common systems in place, often developed not for data management purposes, but rather to ensure common procedures or to lower I/S costs. A typical example can be seen at Blaine Corporation where, although each division has its own products and sales force, common order entry and distribution systems are used by its three major divisions.

As at Blaine, data management may in fact be only a by-product of a common systems effort. However, common systems cannot be developed without surfacing and resolving data definitional disputes, since old systems (and old definitions) will be discontinued. Thus, while the concept of common systems is not new, a number of companies appear to be gaining major data management benefits from implementing common systems and/or from using the standard, sharable data in them.

> The largest division of Waverly Chemicals, a Fortune 500 diversified chemical company, operates about a dozen plants. Since about 1980, the division has emphasized the development of common systems for manufacturing applications such as production scheduling and spare parts inventory. The objectives have been to increase coordination among the plants and to reduce costs in areas such as inventory. A data management group was established as a key to the process.

Several firms which have instituted common systems in a critical business area have found in the era of fourth generation languages, that they have derived significant value from the common data.

> Based on data from its common order entry system, which spans numerous internal manufacturing and distribution divisions, Grand Distributors, Inc., has been able to provide salesmen with complete sales information on a customer by customer basis. This, combined with an estimate of each customer's total purchases from all vendors tells salesmen not only what customers are buying from Grand, but what they are buying from others and, hence, where the marketing possibilities are. This would not be possible if each division had its own order entry systems, since the data would not be compatible. In fact one small division does use its own order entry system, and its data can not be included in the information provided to salesmen.

2.1.3 Information Databases

A third new system "product" is information databases. Whereas the first two products provide data to be used by transaction processing systems and for monitoring real-time operations, information databases are aggregations of data which are primarily used for staff analysis and line management information.

Most "information databases" can be defined as subject area databases for managerial information. In general, for data management purposes, they can be distinguished from subject area databases for operational systems in two primary ways. Information databases do not necessitate rewriting existing systems. Instead, "bridges" are often built from the existing systems to provide the appropriate data to the new subject area database. Depending on the level of compatibility of the existing systems, the bridging process may be straightforward or quite difficult. Also, information databases usually contain aggregated data and, sometimes, external data. [See Note 1.]

> At Windsor Products, corporate management's demands for information led to the development of several new databases. However, existing applications have not been rewritten and the operational databases on which the transaction systems depend remain in place. Automated bridges from the existing systems populate the new "information-only" databases for customer, product and shipment.

> Sierra Energy Corporation has an information database for the key financial information required by the corporate controller's function. This database is incorporated in a system called DMS

(Data Management System). DMS was developed in the late 1970s to improve the flow of financial data from divisions to geographic sectors and, then, to corporate, as well as to increase data integrity and security. A generic bridging mechanism to capitalize on the data collection processes in existing systems was developed, along with download capabilities for forms and report structures. Standardization of data definitions controlled by master tables in a data dictionary helped create the data integrity feature of DMS.

The next case illustrates an effort where an integrated database, originally conceived as an information database, is now also used for transaction processing.

Crockett, the Canadian subsidiary of a major U.S. computer corporation, has developed a single integrated database to support all administrative applications, including accounting and order entry. The effort began as a prototype within I/S. It was originally aimed at removing the operational drain of running report programs on the transaction processing systems, by creating a user friendly managerial information database.

The system which encompasses the database has evolved to include data capture and transaction processing as well as to provide extensive end user access to data. A cornerstone of the new system is an active data dictionary, which is used to determine how to calculate derived data elements and, by end users, as a directory for data definitions. Crockett has devoted considerable hardware resources to the system, but feels it is benefitting from the investment.

2.1.4 Data Access Services

The first three products discussed emphasize developing new databases or files with pertinent, accurate, and consistent data. Some firms in our study, however, focused mainly on improving managerial access to existing data, without attempting to upgrade the quality or structure of the data.

Though the particulars of the efforts differ, each is centered around a small cadre of personnel whose goal is to better understand what data is available in current systems and to put in place mechanisms to deliver this data. These efforts seem to be widely applauded by managers who are able to "get their hands on" existing data. They also have the promise of surfacing questions concerning definitional inconsistencies as managers attempt to make use of the data provided.

> A multi-billion dollar manufacturing firm, Matrac Corporation, has put in place a "data warehouse" for its corporate end users. Briefly, the data warehouse organization is a small group within the corporate I/S organization that provides data access services to users. The group locates data, arranges for extracts, and delivers the data in the user's choice of formats. The group maintains copies of most of the data it provides and arranges for periodic updates. The next step for the data warehouse management team is to develop a data directory designed specifically to assist end users by listing and cross referencing the data available. Included for each data item will be key information such as update frequencies and times, contact people for more detailed interpretation, accuracy characteristics, etc.

Obviously, a key element of delivering data to end users is the provision of both human assistance and improved tools for access and analysis.

> At Blaine Corporation, corporate I/S provides informal data "consulting" services to help end users locate, access, and interpret the data they need.

> There are several information centers at Foothill Computer. Although the centers are charged with the provision of analytical tools and support of end user computing, they also devote significant time to the provision of data consulting services.

Not surprisingly, data access services appear more helpful in companies where data is of reasonable quality than where data is, frankly, a mess. Delivering data in its current form to managers, however, can spur action towards increasing data definition and control mechanisms.

2.1.5 Architectural Foundations for the Future

In most of the firms we studied, managers focused on a limited set of data serving a portion of the corporation. However, there clearly is the danger that by approaching data management in a function by function, business unit by business unit, or subject area by subject area manner, a company leaves itself open to real problems if, in the future, it is desired to integrate data across these boundaries.

To attempt to avoid these future incompatibility problems, some organizations have focused on developing architectural foundations. By architectural foundations we mean policies and standards which, when adhered to, will lead to a well-structured and consistent data environment. Managers allocating resources for architectural foundations are investing in the future without necessarily having immediate benefits in mind. Our cases emphasized two different types of architectural foundations: (1) wide scope strategic data models, and (2) common data definitions.

One view of a firm's data architecture is a corporate-wide strategic data model to serve as an underlying blueprint for all future systems development. James Martin's Strategic Data Modeling approach produces such a model as one of its products [Martin, 1982]. IBM's BSP methodology [IBM, 1981] and Holland's methodologies [Holland, 1983] are others that produce a version of a data architecture. Proponents of these approaches claim that a strategic data model provides an architectural foundation that will lead to consistency of information, more easily integratable systems, and improved productivity in system development and maintenance.

Only one of the ten firms we studied developed a strategic data model primarily for architectural foundation purposes.

> At Waverly's General Polymer Division a strategic data modeling effort was begun for the entire division after the successful implementation of common manufacturing systems. The model has been completed and will be an input to the division's I/S strategic plan. It is still too soon to tell whether that effort will eventually lead to benefits for Waverly.

A second approach to data architecture is the standardization of data definitions. The choice of which data elements should have corporate-wide, standard data definitions is an important architectural issue. In any firm there are some data elements which are so basic to the operation of the business and which are the basis of so much shared communication, that it is critical for all parts of the organization to refer to these elements in the same way. Presumably these data elements should be given global, mandatory definitions. Below the corporate level, it may make sense for a particular division to standardize on certain additional data elements, just within that division. Thus, the standardization of data definitions can be seen as a hierarchical process.

Most of the new system "products" (Sections 2.1.1 to 2.1.3) developed by the firms in our study required that line management and I/S technical personnel agree on the precise definitions of a specific set of data elements, as a prerequisite to building the product. [See Note 2.] In several cases, however, a set of standardized corporate-wide data definitions was intended as an architectural foundation for unspecified future systems development work.

> In 1984, Foothill Computer formed a task force, chaired by corporate I/S, to identify and define key data elements being used in multiple areas of the business. The members of the task force came from the I/S groups in the various functions and divisions. There were no immediate plans to implement the agreed upon definitions. Rather, it was assumed that future systems development work would conform to these definitions. In addition, it was established that any group supplying data to another group within the corporation would be required to deliver that data in conformance with the definitions, if asked. After coming to agreement on definitions for over 200 data elements, the task force has refocused its efforts to concentrate only on those elements for which a specific business impact can be identified and pursued.

> The corporate I/S group at Waverly developed a rigorous methodology for data element definition. As of March, 1985, approximately 150 data elements had been formally defined by the data management staff using this method. Another 500 elements have been identified as candidates for the process. The definitions have been documented on a data dictionary but have not yet been incorporated into applications development work.

Thus, as these examples illustrate, either a wide scope strategic data model or a set of standard definitions can be a product in its own right. Their usefulness can be questioned, however, unless the data model is used to guide future systems development, or the definitions become incorporated in either operational or managerial databases.

2.2 A Variety of Data Planning Processes

We now focus on the planning processes used to select the "products" in the previous section. To many people, DRM is synonymous with a large-scale, strategic data planning and modeling effort. There are, however, other less all-encompassing planning approaches which can be equally, if not more, effective. This section categorizes the planning processes from our case studies into four types: strategic data modeling; "targeting"; "80-20" approaches; and no explicit planning. These approaches represent a continuum of planning processes which assist the organization to identify the target for data management action, and to choose the action (or "product") to pursue. The continuum ranges from well-defined methods (producing relatively detailed, broad scope models of the enterprise and its data), through less formal and rigorous approaches, to no data planning at all.

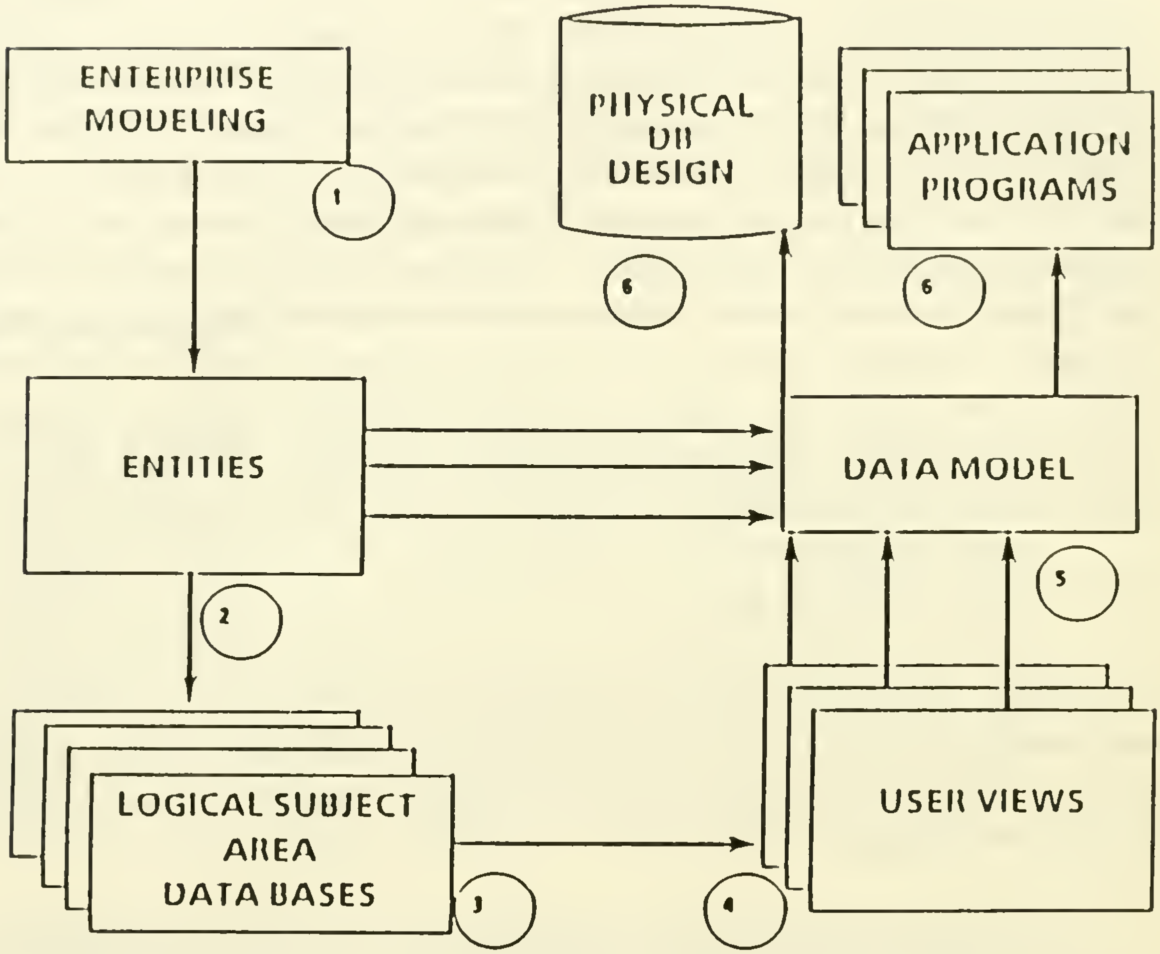
2.2.1 Strategic Data Modeling

Section 2.1.5 discussed strategic data modeling in the context of developing a data architectural foundation. In this section, we primarily focus on using the approach in situations where the near-term objective is to build new systems or databases.

The underlying assumption of the strategic data planning methodologies -- that it is impossible to plan effectively if one does not know what the business is, what it does, and what data it uses -- is difficult to contest. However, of eleven companies we studied, only three used a strategic data modeling approach, and as yet only one of those has succeeded in implementing that model in actual systems. Informal discussions with many other firms have surfaced many disappointments with such approaches, and few successes.

The diagram in Figure 4, adapted from James Martin's Strategic Data-Planning Methodologies, is representative of several data-oriented strategic planning approaches. The left side of the diagram shows the top-down planning approach, leading to the bottom-up design shown on the right side of the diagram.

Figure 4. Strategic Data Planning



Adapted from Figure 7.3, Strategic Data-Planning Methodologies (James Martin)

The process begins with the development of an enterprise or business model, (Box 1 in Figure 4). The enterprise model depicts the functional areas of the firm, and the processes that are necessary to run the business. The next step is to identify corporate data entities and to link them to processes or activities, (Box 2). Data requirements are thus mapped onto the enterprise model, leading to the identification of subject areas for which databases need to be implemented, (Box 3).

In general only selected portions of the enterprise model and subject area databases are selected for bottom-up design. Building the logical data model is the first step. The data model, (Box 5), results from a synthesis of detailed end user and management data views, (Box 4), with the results of the previous top-down entity analysis, (Box 2). Database design and subsequent design of application programs, (Boxes 6), proceed from the logical data model.

As mentioned in Section 2.1.5, Waverly Chemicals has successfully developed a strategic data model for use as an architectural foundation for future systems, but has not as yet implemented any systems using the data model. Dobbs Insurance is so far the only reasonably successful example in our study of this approach leading to actual systems development.

> Dobbs did successfully use the top-down bottom-up approach in its group insurance division, modifying the approach somewhat to reduce the detail required. Within a few years, the division rewrote virtually all its systems to conform to the data model it had developed. Some departures from the model have been necessitated on occasion to meet the requirements of short-term deliverables. Efforts to resolve these "nonconforming" systems with the data model are made at a later point in time.

While the strategic data modeling approach resulted in new systems at Dobbs, it did not at Foothill Computer.

> In the late 1970s, Foothill Computer recognized data management as an important issue. A group, composed primarily of senior I/S professionals, began to study and then to develop top-down, data-oriented methods for strategic systems planning. After much exploration, the group developed a strategic data planning method that integrated both data-oriented and process-oriented views of design.

> In 1982, the method was used to analyze one major functional area, the flow of orders through the company. The project team worked for six months to define and get consensus from the worldwide business units on a logical data model. They also reached agreement on data elements and definitions. Logical subject area databases were designed, and much effort was spent validating those databases with user views. The systems, however, were never implemented. A host of problems led to discontinuing the project. These included the difficulty of consolidating disparate user views, doubts about whether all the key user views had been obtained, and not least of all, a multi-million dollar price tag for implementation. The course of action finally chosen was to make extensive modifications to an existing system to address its shortcomings.

There were 14 other data management efforts in the eleven firms we studied. In none of these did the planning and implementation process proceed as suggested by the strategic data modeling methodology. Instead, we saw a variety of other approaches used successfully. Figure 4, discussed previously, depicts a generalized strategic data planning process. Figure 5 illustrates the actual planning process in four of the case study sites.

> As shown in Figure 5a and 5b, both Spectrum Electronics and Waverly Chemical chose to develop manufacturing systems without having used any top-down, data-oriented planning methodology to arrive at that decision. They started with a particular set of applications (1), then developed a logical data model (2), and finally designed the physical database and the applications (3).
>
> Sierra Energy, (Figure 5c), started with a particular user's view (1)--that of the corporate controller--and designed a data model (2) and physical data bases (3) from that perspective. Methods for "bridging" data from existing applications were then developed (4).
>
> At Windsor Products, (Figure 5d), the data management group in corporate I/S consulted extensively with the distributed I/S groups and with senior management to reach a consensus on what new corporate-wide databases would have the greatest impact on the business (1). The data management group then consulted with all users of that data as it built the logical data models for those databases (2). They essentially replaced the entire left side of the diagram by wide-ranging deliberation within the distributed I/S organization, then proceeded to build databases and "bridges" (3).

In all of these cases, the firms either skipped or abbreviated the "left side" or top down portion of the top-down planning bottom-up design

Figure 5. The Planning Reality

Figure 5a.

SPECTRUM

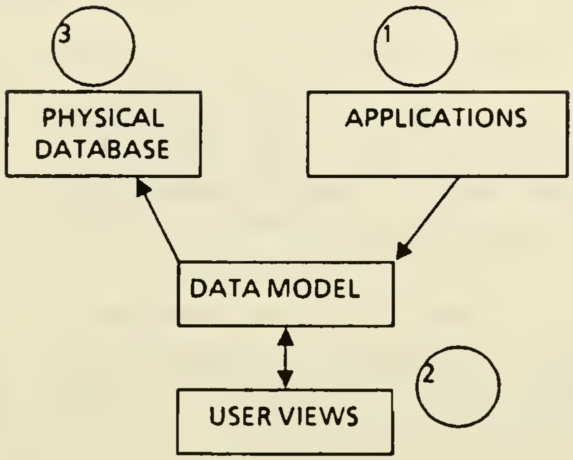


Figure 5b.

WAVERLY
(General Ploymer Div.)

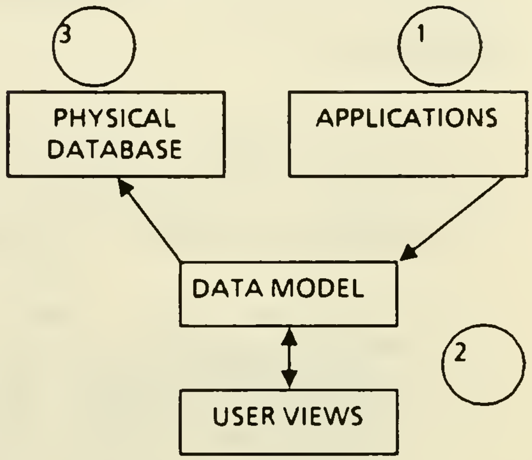


Figure 5c.

SIERRA

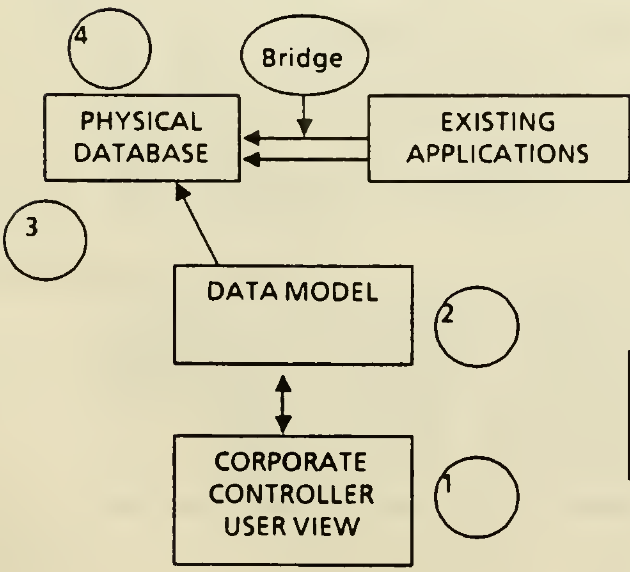
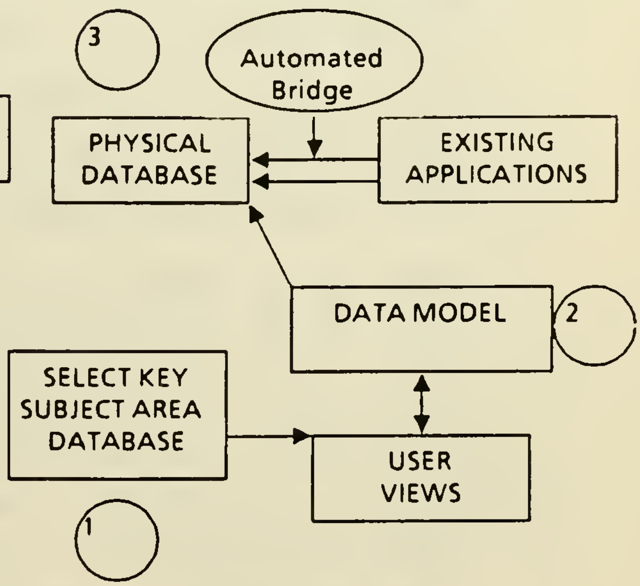


Figure 5d.

WINDSOR PRODUCTS



process. In fact, they followed the alternate planning processes to be discussed in Sections 2.2.2 - 2.2.4.

Some companies, like Foothill, that do follow a top-down, strategic data planning process have significant difficulty implementing the plan. Upon reflection there appear to be a number of reasons which help explain why more firms are not successfully using strategic data planning methods. Among them are:

- The strategic data planning process, done in detail and with a wide scope, can be very time consuming and expensive. Also, for the process to be successful, key managers must commit significant time and effort. This commitment is often difficult to obtain (and keep) from these busy individuals.

- It is not always clear to the planners or top management whether strategic data modeling is being done to create an architecture or to immediately build new systems. It is difficult to manage the expectations of those involved regarding the results and benefits of the process.

- Total implementation of a wide scope data modeling effort can be extremely expensive. There is a tendency to avoid these new costs, especially if many of the existing systems which represent a huge investment are still effective.

- When implementing only a subset of the plan, it can be difficult to bridge the gap from the top-down plan to bottom-up design. If proposed and existing systems interface along different boundaries, it may be hard to isolate and replace a subset of existing systems with a subset of proposed systems. The use of application packages also creates interface and boundary problems.

- Because of the upfront effort needed, organizations face a longer and more expensive development process for the initial systems developed with this planning method. Line managers do not like to see project schedules lengthened. Similarly, I/S managers, who have incentives to deliver quickly and contain costs, may resist the additional effort involved.

- Often the business will change while the plan is being developed and implemented.

- The methodology requires new I/S skills.

Thus, for all these reasons, it is difficult to gain commitment to the process and manage expectations regarding the outcome of strategic data planning. We will return to some of these items in Section 4.0 which discusses several fundamental managerial issues affecting the implementation of data management efforts in general.

2.2.2 Targeting High Impact Areas

Most corporations that skip or abbreviate the top-down planning methodology do not act without a plan. There are a variety of alternative planning processes which can be used.

The most common process is the "targeting" of a particular function or other business area. In some companies important problem or opportunity areas are quite clear without an extensive analysis.

> In Sierra Energy, the controller knew from his experience that he needed, (and did not have), consistent, accurate data for corporate reporting purposes. At Windsor Products, it was quite clear that consistent and accurate customer data and product data needed to be shared to support changing business demands and, after some discussion, it became clear that shipment data was also critically needed. At Spectrum Electronics, the head of the I/S department was certain, and was able to convince the manufacturing vice president, that an integrated manufacturing database and common systems were the answer to coordination problems and high systems costs.

In none of these companies was a rigorous data planning method used. But, in all, there were key managers who could visualize the benefits of improved data quality. In each case, a data management program was undertaken which was limited in scope, but was effective and implementable.

2.2.3 "80/20" Planning Methods

In some firms, there is a desire to get the major benefits of global data planning without having to invest the amounts of time and money necessary to carry out a full-scale strategic data planning process. The aim in these cases is to zero in quickly on the key "products" to be implemented (bottom-up), while reducing the amount of effort spent in a global planning (top-down) phase. This type of planning can be termed an

"80/20" approach, after the adage that for many undertakings, 80 percent of the benefits can be achieved with 20 percent of the total work.

> Windsor Products, having carried out its first round of data management in a quick targeted manner (as shown in Figure 5d), found itself uncertain as to the next subject databases to implement. Management decided, however, that a full strategic data model was not necessary for its purposes. What it needed from a planning process was a means to fairly rapidly identify the next round of candidates for subject area databases. It therefore developed its own abbreviated planning approach which it calls "Data Charting". This involves identification of the major data aggregates in the corporation and the groups which used them. The use of this method took less than six man months of effort, and involved reviews by line managers in all parts of the business. This "Data Chart" is serving as the basis of planning for the next round of information databases.

A different type of "80/20" approach was used to identify key data elements that should have common corporate-wide data definitions at Foothill Computer.

> In 1984, Foothill put priority on developing common data definitions to promote data consistency, as described in Section 2.1.5. A Key Data Task Force was formed to identify and resolve high impact data definitional problems. The method used was to ask distributed I/S groups to each nominate 50 important, cross functional data elements. There was a 30 to 40 percent overlap among the twelve lists of data elements nominated. Work then proceeded on the "product" of developing common definitions for the identified critical data elements.

The problem with strategic data planning approaches is the investment of time and dollars needed to obtain results. Drawbacks of "targeting" include the probable inconsistencies that will arise from multiple targeted projects, and the fact that, in some companies, the most appropriate targets may not be evident. An "80/20" approach, while not providing the detail of strategic data planning or the quick hit of targeted approaches, does offer the major benefits of both previous approaches.

2.2.4 No Planning Process

There are also data management actions that can be taken without doing any data-oriented planning. For example, if a decision is made to provide better access to existing data, without addressing changes in the form of that data, then no data planning methodology is needed. This was the case at Matrac with its data warehouse approach, described in Section 2.1.4.

2.3 Bounded Scope

In our study, no firm attempted to manage all the data used by the corporation; all limited the focus of the effort in one or more ways. An important factor in the success of data management efforts is that the scope of the effort be carefully selected. As described in this section, scope defines which part of the organization is to be included in the data management effort.

Most firms focused on a functional area, such as Spectrum Electronics which targeted manufacturing (see Section 2.1.1), or Sierra Energy which targeted corporate finance (see Section 2.1.3). The scope of some data management efforts was divisional. An example is the case of the Group Insurance Division of Dobbs, described in Section 2.1.1.

In some of our cases, the scope of the planning effort was corporate. However, in all but two of these instances, the product was a relatively small set of standard data definitions. Windsor Products focused on a selection of corporate-wide subject areas: customer, product, and shipment data. Matrac focused on making data available to the corporate headquarters staff. Thus, these efforts, while corporate in scope, were limited to a small subset of the total data used by the corporation.

In addition to functions and divisions, other suborganizations exist within a corporation. Among these are: groups, sectors, geographic districts, and product lines. Some organizations may choose one or another of these units as a locus for data management efforts. As the next section will show, the scope of the data management effort is substantially determined by the business objective.

2.4 Business Objectives, Not Conceptual Justifications

The proponents of data management far too often base their arguments on either the conceptual soundness of viewing data as a resource or the rationale underlying data-centered systems design. They assert that processes change while data is relatively stable, and therefore data should be the key element in I/S planning. Or, they argue that global data management is essential because one needs a global plan before developing the individual pieces. Without such a plan, the pieces built one at a time will not fit together, and the result will be entrenched inconsistencies in myopically designed systems. While these arguments are appealing, they do not engender action in the pragmatic, cost-conscious world of the business manager.

The successful data management processes we observed have been aimed, for the most part, at solving a clear and specific business problem or exploiting an opportunity. This section discusses the kinds of business problems and opportunities which motivate executives to consider more intensive management of the data resource.

2.4.1 Coordination within or across Business Units

A major motivation for data management action is the perceived need for better coordination of activities, either within specific functions or business units, or across them. Improved coordination requires an enhanced ability to communicate within or between organizational units. In practical terms this means more than the rapid sharing of data. In order for data to be shared usefully, it must be in a language understood by all involved. This implies common data definitions and common codes for key data elements.

There are a number of clear examples of coordination as a motivation in our case studies.

> Both Spectrum Electronics and Waverly Chemical felt the need to standardize their manufacturing systems so that many plants could be coordinated more effectively. In both these cases,

their efforts have led to significant benefits. Better coordination between the production schedules of separate plants has led to reduced in-process and inter-plant inventories; the coordination of spare parts inventories has led to smaller inventories and reduced downtimes; and coordinated purchasing operations has brought economies of scale and special arrangements with vendors.

Sierra Energy is a different type of example. In order to meet external reporting requirements, Sierra needed to better coordinate the financial information being fed to corporate from diverse decentralized divisions. Again the need was for a common "language" so that critical information could be rapidly and accurately available to the headquarters units.

2.4.2 Organizational Flexibility

A second category of motivation for data management is the desire for greater organizational flexibility to allow either an internal restructuring of the organization, or a refocusing of the organization due to changes in the environment.

Waverly Chemical merged two large manufacturing divisions in the mid 1970s, and faced major problems with accounting systems which had been designed and implemented separately in the original divisions. The company faced similar problems in the late 1970s when it again reorganized, this time combining five old divisions into two new ones. It was quite clear to senior management that there would be other reorganizations in future years. As long as each division had its own accounting systems with much incompatible data, the problems would persist.

In regards to refocusing, a number of companies we studied have been faced with important changes in the marketplace which have put intense competitive pressures on them to change from a product focus to a market or customer focus.

At Dobbs, each product line in the Group Insurance Division sold directly to customers and maintained its own customer information. Each product line had its own set of customer codes. As the competitive environment changed, the inability to present a single face to a customer interested in many types of services put Dobbs at a disadvantage.

Organizational flexibility is often hindered by data structures which have been designed to support particular applications or suborganizations, but which are not flexible enough to be able to provide new strategically important "views" of the business.

2.4.3 Improved Information for Managers

A third motivation for more effective data management is the need for improved information for senior managers, middle management, and key staff personnel. These information consumers want two things: improved access to data and improved data quality.

> Top management at Windsor asked for a list of their fifty largest customers. They were told that because the data was organized around product lines with different customer codes within each product line, it was very difficult to respond to their request.
>
> At one bank, the quality of computerized management information on such items as loan position and bankers acceptances was so poor that managers relied on manually prepared, redundant reports. Since it took almost three months before reported errors were corrected, some managers stopped notifying systems personnel about errors and only used the manual reports.

When problems such as these become important enough to management, strong motivations arise to improve data quality and access.

2.4.4 I/S Efficiency

As backlogs increase, and as I/S costs climb, there is a clear need to adopt methods and tools that show promise of increasing development productivity or reducing maintenance costs. Improved data management is considered to be a way of addressing both of these problems. In a more standard and stabilized data environment, programmers should be able to avoid having to constantly "reinvent the wheel" of data structure and definitions. Reduced data redundancy should also trim costs. Presumably, reduced inconsistencies will make it easier to access information for either new applications or new reports.

In few of the cases we studied was I/S efficiency a primary motivation for taking action. To the contrary, a number of firms felt that the data management actions they were undertaking would involve greater, not lower, I/S costs in the short run.

Some firms did look for improvements in I/S efficiency as a secondary benefit, and some have achieved it.

> Spectrum eliminated separate programs and programming staffs in its plants by centralizing all data processing and using a centralized database and common software for all manufacturing systems. The company also claims to have reduced maintenance costs by eliminating not just redundant data, but the programs that updated redundant data, and the programmers who maintained those programs.
>
> Crockett claims a 40% reduction in development costs through use of its active data dictionary, the refocusing of the development process on eliciting business rules rather than programming procedures, and user generation of their own reports.

2.4.5 Summary of Motivations

What all these motivations have in common is that they are tightly linked to the business. They reflect competitive, managerial, and cost factors. None of them draws on the conceptual justification that data should be considered a corporate resource. The companies participating in our research are taking steps to better manage data for concrete business reasons.

In a few cases, these motivations to better manage data emerged from a line manager's need to respond to changes in his or her environment. I/S departments, however, have also had a key role in initiating many of the successful efforts we have seen. Many line-sponsored efforts have occurred after months, or even years, of education by I/S managers.

3.0 FOUR CRITICAL COMPONENTS FOR DATA MANAGEMENT: SUMMARIZING THE CASES

The four critical components for data management action, presented previously in Figure 3, are: business objectives, scope, data planning process, and "product". These components provide a rough framework for thinking about data management actions in a corporation. In Section 2, we discussed the choices for each area separately. In this section the framework is presented together with the actual choices made by the case study firms. A few evident patterns emerge.

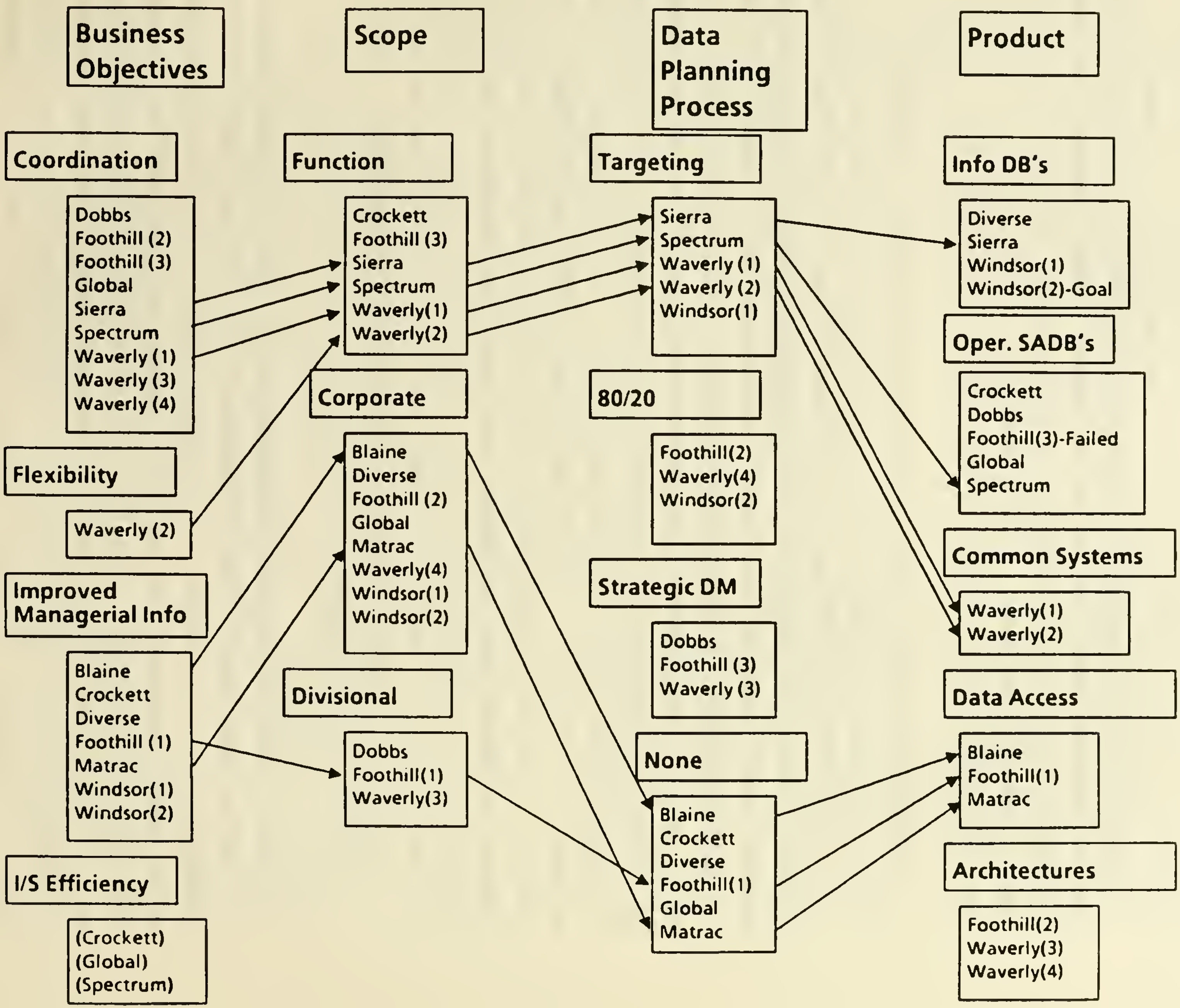
Figure 6 shows the filled in framework. In each column the firm's name appears under the alternative it used in a major data management effort. Where a firm had more than one major effort, its name appears several times, with a subscript to distinguish between the efforts. Refer back to Figure 1 for capsule descriptions of each effort.

Several observations are evident from the figure. First, there is one especially prominent path through the diagram. This path, with 4 companies, is from a coordination or flexibility business objective, to a targeted planning process, to a functional scope, to a variety of products. The path reflects the existence of managers in specific functional areas who have been able to see a positive benefit/cost ratio in their area of responsibility.

Second, another path that stands out (with 3 companies involved) is from an objective of improved management information, with various scopes, to no planning process, to a "product" of improved data access. This is an important option, with real benefits at minimal cost.

In addition to these two prominent paths, Figure 6 shows a rich array of choices, guided predominantly by the particular business objective of the firm, and its unique situation. Strategic data modeling is only one of several appropriate planning processes. Likewise, subject area databases are not the only "product" through which benefits can be delivered. Common application files provide useful, usable data; information databases

Figure 6. DM Framework With Companies



"bridged" from existing application systems contain useful data for managers; and even the improvment of data access services can have important benefits.

There is no single path to improving the management of data in a corporation. Multiple options exist. The correct choices for any organization are contingent upon the readiness and needs of that organization at a particular point in time.

4.0 FUNDAMENTAL MANAGERIAL ISSUES AFFECTING DATA MANAGEMENT IMPLEMENTATION

Our interviews with I/S and line managers, and our own reflection on the concerns raised, have suggested five fundamental managerial issues which seem to underlie many of the specific problems affecting the ability to effectively implement data management efforts. These are (1) the trade-off between short-term deliverables versus long-term plans or architectures; (2) the tendency to centralization inherent in data management; (3) the need for new organizational roles and responsibilities; (4) the impact of data management on I/S culture; and (5) the process of effective introduction of innovations into an organization. In this section we present our early thoughts on these issues as a means of laying the groundwork for future research.

4.1 The Trade-off between Short-term Deliverables versus Long-term Architectural Foundations

Managers considering data management actions today are faced with a trade-off between short-term deliverables and long-term architectural foundations. The development of a data architecture is in essence an infrastructure decision, like the building of roads. The benefits are long term and often diffuse. A fact of life in American business is the short-term focus of line managers. Decisions to allocate resources to data architecture are in direct conflict with managers' focus on quickly demonstrable results, near-term earnings per share, and this year's return on investment. Not surprisingly, most of the successful efforts reported in this paper have tended toward the short term end of the spectrum.

The danger of limiting the scope of data management action to areas with a near-term payoff is that this may result in problems later on if the local solutions ultimately prove incompatible with each other. Moving to the "architecture" end of the continuum, however, also has its problems. Efforts to prevent data inconsistencies by developing corporate-wide strategic data models are usually very expensive, and many have not succeeded.

What is needed is a less extensive, less expensive approach to data architecture than strategic data modeling. In order to develop such a new approach, we need to understand better what aspects of current data design might conflict with future needs, and what principles can guide current efforts in constructive, minimally restrictive ways. The "80/20" data planning effort at Windsor Products seems a very useful move in the right direction.

4.2 The Centralizing Tendency of Data Management

Underlying any effort toward more effective data management in an organization is the reality that improved data management can lead to greater centralization of decision making. Increased standardization of data does facilitate increased central control.

For example, standard data definitions may be established as common systems are developed. If the resulting data is made accessible to senior executives, they may have a vastly enhanced ability to compare operational details of business units under their jurisdiction. There is a tendency to act on this data. In fact, several of the systems described here were instituted to facilitate "central" coordination and control. Even if increased centralization is not a design objective, it may be a result.

Data management thus presents corporations with a familiar dilemma. For a particular business unit in its unique environment, there is a need to make explicit choices concerning the balance of centralization versus local flexibility that will best serve the business as a whole. Of equal, if not greater, importance is the recognition of the almost inevitable resistance to the implementation of actions which shift real or perceived power in the organization [Markus, 1983].

4.3 New Organizational Responsibilities

Data management is changing old and creating new responsibilities in organizations. It is apparent that new organizational units are being established to handle various data-related activities. In addition, policies are being set that describe the responsibilities of personnel with respect to areas such as data collection, access, and security. Perhaps most significantly, the new responsibilities have an impact on not only the I/S department, but also the roles of line managers.

In addition to database administration groups and data administration groups, which are highlighted in the literature [Ross, 1981] [Gillenson, 1985], some companies in our study had "corporate DRM groups" and "data access groups". A few firms had created corporate DRM planning groups, responsible for the development of corporate plans, guidelines, and, sometimes, architectural foundations. The emergence of data access groups as an important part of the organization supporting end user computing deserves to be emphasized. For the end user, assistance with hardware and software is not enough. Increasingly, end user support includes data consulting. (See Section 2.1.4.)

Establishing the appropriate organizational units is just one part of the organizational redesign effort. To achieve effective implementation, responsibilities must also be assigned for establishing and executing administrative policies and procedures related to data quality. The roles of data owners, data suppliers, data custodians, and data users need to be clarified.

> One example of how the clear assignment of administrative responsibility helped data quality comes from Diverse Conglomerate, where an extensive information database used by the chief executive was fed by many operational systems from many divisions. The quality and consistency of the CEO's database depended upon coordinating the update data from the divisions so that the database reflected a view of the corporation at a single point in time. In order to manage what was becoming a difficult problem, the group in charge of the CEO's database developed a "master data input calendar" for critical data files coming from the divisions. The MDIC specified the schedule of updates from each division, and the person responsible for data arriving on time. Thus it formalized the operational coordination problem.

> The Key Data Task Force at Foothill Computer, mentioned in Section 2.1.5, has issued standard definitions for some 200 data elements. As part of the process, the task force identified a "controller" for each data element. The "controller" of a data element is responsible for its definition, but not necessarily for the acquisition, updating, or accuracy of the physical data. The typical "controller" is a senior business manager. Not all of the managers selected as data controllers readily took on the responsibility.

At this point in time, the organizational responsibilities with respect to data management are still evolving. Significant thought must be given to the establishment of appropriate roles and organizational units. In many cases, these fledgling groups must be given strong sponsorship, and be protected from the vicissitudes of corporate politics.

4.4 Impact on the I/S Culture

A fourth managerial issue is the impact of data management on the I/S culture. DRM implies significant changes in the ways systems planning and design are carried out. As Durrell [1985] points out, "data administration challenges the basic process-oriented approach that has been employed during the last 20 years. This can be disconcerting and sometimes insulting to many long-time DP professionals" (p.ID/29).

> The new data-centered system used at Crockett (Section 2.1.3), has had a major impact on the work of programmers and analysts. The major activity now of these people is working with business managers to define the business rules governing the meaning of data elements, and entering those rules into the data dictionary. There is really no such thing as an applications

> programmer at Crockett anymore. Not surprisingly, this has had a significant impact on the attitudes and turnover of the programmer/analyst staff there. A great many programmers found their skills of no value in the new environment, and many left during the transition period. While the level of excitement of those who remained is high, Crockett personnel admit that they have more luck hiring recent graduates than they do bringing in experienced programmers to fill the new applications development role.

The problem is not only one of teaching information systems professionals new skills. There must also be organizational changes to support the move toward data-oriented design. For example, current incentive schemes reward programmers for implementing systems on schedule and not for conforming to, or petitioning for changes to, the data model or data standards. Without changes to these incentive structures, when system deadline pressures become high, programmers have a strong tendency to develop their own local data structures rather that enter into time-consuming negotiations with the data administrator for changes to the corporate data model.

The cultural and organizational changes implied by adopting a data-oriented approach to systems development are important to anticipate. The incentive structure appropriate for programmers and the skills needed by them are two obvious areas. Other important changes will probably become apparent as organizations gain more experience in this area. The adoption of new methods and tools reflecting a major change to the traditional approach to systems development will not be successful if the organizational change is not carefully managed.

4.5 The Process of Effectively Introducing Innovations into the Organization

The implementation of initial data management efforts can be usefully viewed as the process of effectively introducing innovations, i.e., new methods or tools of unproven value, into the organization. Rogers suggests that in general the diffusion of innovations is dependent, among other things, on five characteristics of the innovation. [Rogers, 1962] These are (1) the relative advantage of the innovation over its alternatives; (2) the observability of the results; (3) the compatibility of the innovation with existing values, past experience, and perceived needs; (4) the complexity of the innovation; and (5) its trialability, or the extent to which the innovation can be experimented with on a small scale, low risk basis.

Rogers' theoretical argument can help explain why it has been difficult to implement data management actions in many corporations. First of all the relative advantage of most data management actions compared to current practice is not known. To date, in most organizations, there have been few results to observe. Where results are available, it is extremely hard to separate, and quantify, the impact of data management from other related (or unrelated) actions.

As noted in Section 4.4, a data-focused approach to I/S is not compatible with the existing application/procedure/project focus, where the goal is to build individual systems to specification on time, rather than to create a data architecture to meet current and future needs. Certainly, data management involves a great deal of complexity as the walls between applications are torn down and interrelationships between systems, functions, and organizational units are examined. Finally, very often data management actions have not been presented as trialable, small scale, low risk efforts. Instead, DRM has been sold on the basis that a major financial investment and top to bottom commitment in the organization will be needed to achieve results.

If one does adopt Roger's perspective on the diffusion of innovation, it is easy to understand the difficulties that organizations have encountered in implementing large-scale data management efforts. This perspective argues for more limited approaches where the relative advantage is clearer, the impacts more observable, and where complexity is lessened. In addition, limited approaches can be viewed as trials or experiments. When the first trial is successful, the organization will probably be ready for a more ambitious second trial, and ultimately for significant investment in "data infrastructure".

5.0 CONCLUSION

Three major themes emerge from our study. First, it is clear from our research that managing data is a significant issue in corporations today. The ability of firms to access managerially relevant data, to coordinate operations, to reorganize, or to change their strategic focus can be severely limited by poorly managed data.

Second, there is no one, clear-cut approach to managing data. A wide range of options exist that can be tailored to the needs of a particular business. The appropriate planning process to use and "product" to deliver depend heavily on the particular business objective and organizational scope. As demonstrated by our case examples, there are many contingencies to be considered in data management and, thus, multiple paths to take. Another view of these contingencies is that data management actions can have an impact on three aspects of a corporation's data: its infrastructure, its content, or its delivery.

By infrastructure we mean design or definitional standards which limit current local flexibility in favor of greater global flexibility in the future. Enforced definitional and coding standards are one example. The development of common systems, with their implicit enforcement of a single set of data definitions, is another of the many data management actions which can contribute to a data infrastructure. Actions to build a data infrastructure tend to be both difficult and expensive, and the benefits from such actions are most often realized only in the longer term.

Content refers to the choice of what data to maintain, and also to policies that address the accuracy of that data. Systems to capture new or more detailed data, and decisions to purchase external data are examples of actions which affect data content. These efforts tend to be moderately expensive, with benefits in the middle term.

Delivery refers to making existing data available to managers who need it. Data consulting services, extract policies, and the provision of fourth generation reporting tools are examples of mechanisms to improve delivery. Actions in this area tend to be less expensive, and have short term benefits.

The choice of where a firm should best allocate its resources - between infrastructure, content, and delivery - depends very much on the willingness and ability of senior management to invest now for future benefits. Current systems, management style, resource constraints, strategic vision, and the evolution of the industry are all important factors.

The third theme is that no matter which path is chosen, there are a number of managerial and organizational issues which must be addressed in order to successfully implement data management efforts. The managerial affinity for short-term results, the centralizing tendency of data management, the need for new organizational roles and responsibilities, and the impact of data management on the I/S culture all present issues that, if not managed well, will severely inhibit the effectiveness of data management efforts. Treating initial forays into data management as a process of introducing innovation can help managers understand which efforts are practical in their organizational environment.

NOTES

NOTE 1.

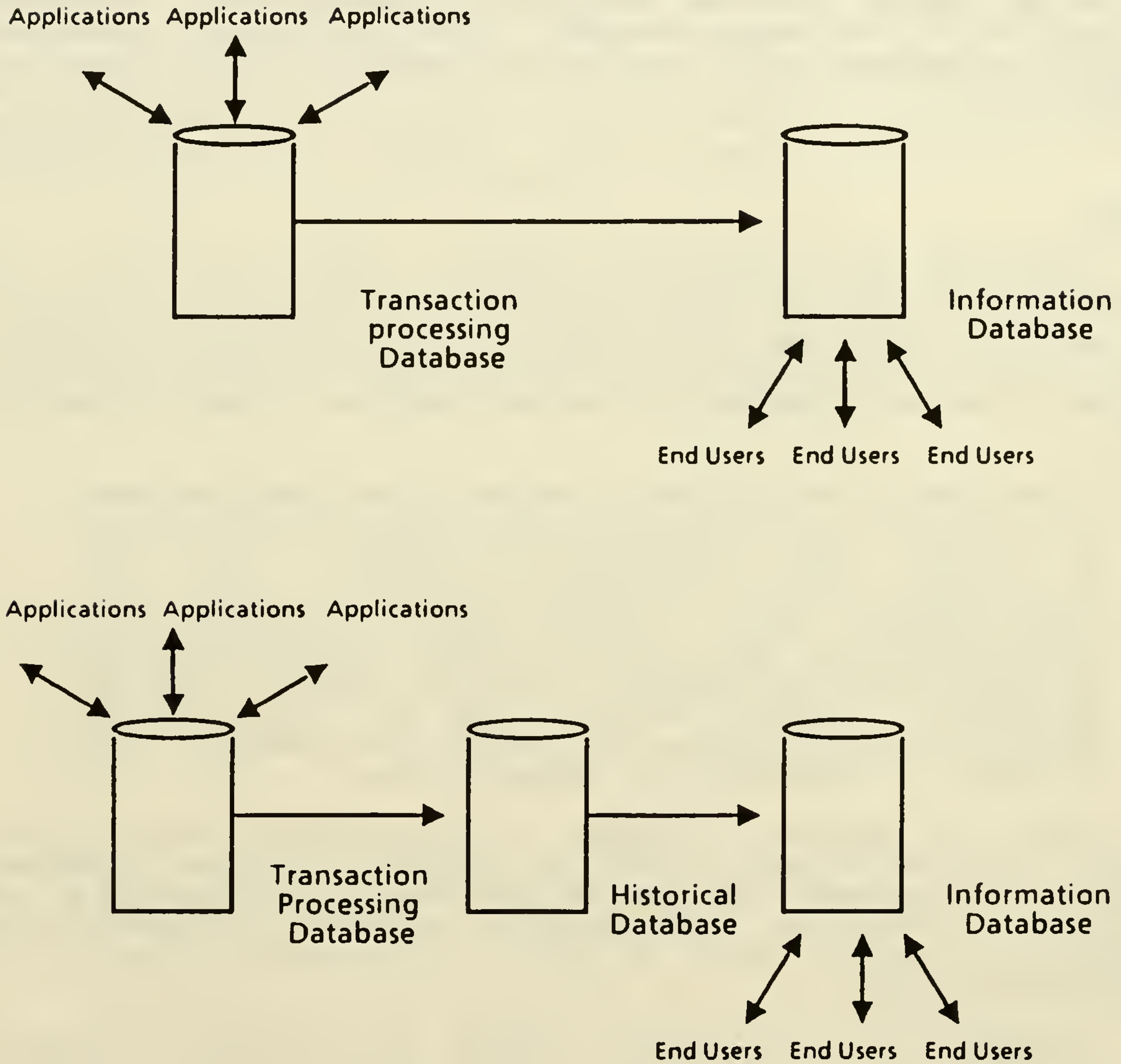
The two diagrams in Figure 7 show alternate visions of the relationship between operational databases and information databases. In the first diagram, transaction processing (TP) systems are segregated from end user information databases to allow better tuning of the high volume TP applications and databases. Only selected, and usually aggregated, data is downloaded to the information database, making the latter smaller and more easily tuned to the very different load and demand pattern of end user queries. Thus, the TP database has complete data, but only for a limited time horizon. The information database has only selected and aggregated data, but for a longer time horizon.

This two block model requires either that we know in advance what detailed historical data managers will be interested in, or that we download to the information database all detailed data purged from the TP database, or that we accept that some future questions will have to go unanswered.

The three block model, pictured in the second diagram in Figure 7, recognizes that the questions of interest to managers querying the information database will change over time, in ways that cannot be predicted in advance. Therefore, it adds the historical database which is intended to carry all data about company transactions at the lowest level of detail available. This is obviously a large database but one with very lenient efficiency demands. Extract programs to feed the information database can change as needs change, and if it is necessary to view data in a new configuration, both current and historical data can be put into that form.

> An extreme example of problems with historical data comes from Southern Cross Products, which sells consumer and gift products. There is currently a major push at Southern Cross to develop better product line sales forecasts by including economic indicators in the anlysis. Thus, the marketing analysts want data on past sales to as far back as 1950. Unfortunately, the data processing group maintains data for one year on tape, then discards it. The solution has been to locate "old timers" who have been with the company for years and might have old printouts of sales data hanging around on a shelf somewhere. The data which the marketing analysts have found in this manner has been rekeyed to create a patchwork database for the forecasting analysis.
>
> A more typical example comes from Boston Mutual. In this insurance company, a front-end policy preparation system records all agents who share in the credit (and compensation) for any particular policy. When the policy actually takes force, however, the preparation information passes to an information

Figure 7. Conceptual Views of Operational and Informational Data



database which carries only the first agent listed. Recent business pressures have forced Boston Mutual to begin to dock agents for policies which are not maintained beyond the first year, but the data to do this accurately is not available.

The two models in Figure 7 illustrate an important architectural choice that firms must make. Depending on the business demand for data and on technical constraints, firms may make different choices for different categories of data. For instance, detailed sales-related data may be captured using the three block model with an historical database because of the likelihood of having to reaggregate in different forms as marketing structures change. Detailed manufacturing data requiring real time accuracy, but with no presumed historical significance, might be purged from the transaction processing database without loading it into the historical database.

NOTE 2.

Establishing data definitions was important in many of our cases. Data definitions and codes are key standards for a consistent data environment. The process of resolving specific data definitions, however, is not easy. While as many as 50 percent of the data elements in a given area may be defined without any disagreement, those that remain may be very difficult to resolve.

As mentioned in the section on common systems, data management began at Waverly's General Polymer Division as part of an effort to develop common manufacturing systems. Today, the standardization of data definitions is viewed as central to GPD's data management program. The cornerstone of the program is an internally-developed data dictionary, with over 5,000 total and 500 standardized data elements. A committee composed of the data administrator, systems analysts, and user representatives facilitates the definitions of standard data elements. These definitions are then used by all new applications.

Many of the companies we studied, including Waverly mentioned above, used a task force of both I/S and user managers to resolve definitional problems. I/S personnel are familiar with the technical aspects of the definition, including how implementing the definition might affect other systems. User managers can see the impact on their business practices from different definitions.

A technique we saw at Foothill Computer for resolving definitional disputes is to refocus the dispute from "what should the definition be" to "who has final authority for deciding". This person or position is then the "owner" of the data element, and has both the right to decide on the definition, and the responsibility of responding to others who are unhappy with the current definition.

Metadata standards add structure to the definition process by explicitly setting out what a data definition must include to be complete. Several of the companies in our study had formal metadata standards to improve the quality of data definitions and the productivity of the definition process [Symons and Tijsma, 1982]. These standards ranged from being data-type classification schemes to being format "languages" with syntax rules for the structure and completeness of definitions and a controlled vocabulary of permitted terms to be used in definitions.

In spite of the importance of developing and managing standard data definitions, tools in this area are not yet technically adequate. Most data dictionaries on the market were designed to support a particular software environment, and thus are extremely cumbersome or unworkable in the typical corporation with data on multiple hardware and software systems. This has forced several of the corporations we studied to invest in designing their own data dictionaries to be able to meet their needs. Satisfaction of the corporations with their proprietary data dictionaries varies, but it is clear that there is an unfulfilled need here.

REFERENCES

Appleton, Daniel S. "Business Rules: The Missing Link", Datamation, October 1984.

________ "Information Asset Management", Datamation, February 1, 1986.

Chen, P.P.S. "The Entity-Relationship Model - Toward a Unified View of Data", ACM Transactions on Database Systems, Vol. 1, No. 1, March 1976.

Coulson, Christopher J. "People Just Aren't Using Data Dictionaries", Computerworld, August 1982.

Diebold, John "Information Resource Management--The New Challenge", Infosystems, June 1979.

Durrell, William "The Politics of Data", Computerworld, September 9, 1985.

Edelman, Franz "The Management of Information Resources: A Challenge for American Business", MIS Quarterly, Vol. 5, No. 1, March 1981.

Gillenson, Mark L. "Trends in Data Administration", MIS Quarterly, Vol. 9, No. 4, December 1985.

Holland, R.H. "Tools for Information Resource Management" Paper presented at The Guide Conference, New Orleans, Louisiana, November 9, 1983.

Huber, George P. "The Nature and Design of Post-Industrial Organizations", Management Science, Vol. 30, No. 8, August 1984.

IBM, Business Systems Planning, IBM Manual #GE20-0527-3, July 1981.

Kahn, Beverly, K. "Some Realities of Data Administration", Communications of the ACM, Vol. 26, No. 10, October 1983.

Markus, M. Lynne "Power, Politics, and MIS Implementation", Communications of the ACM, Vol. 26, No. 6, June 1983.

Martin, James Strategic Data-Planning Methodologies, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1982.

Rogers, Everett M. Diffusion of Innovations, Free Press, New York, 1962.

Ross, Ronald G. Data Dictionaries and Standard Data Definitions: Concepts and Practices for Data Resource Management, New York, Amacon, 1981

Symons, C.R. and Tijsma P. "A Systematic and Practical Approach to the Definition of Data", The Computer Journal (British Computer Society), Vol. 25, No. 4, November 1982.

Tillman, George D. "Data Administration vs. Data Base Administration - There Is a Difference", Infosystems, February 1984.

Voell, Ronald F. "The Evolution of Data Administration", Text of an address delivered at the Information Management Conference, New York, October 17, 1979.

Zmud, Robert W. Information Systems in Organizations, Scott, Foresman and Company, Glenview, Illinois, 1983.

Date Due

NOV. 29 1992		
MAR. 19 1993		